

University of Groningen

Incomplete datasets obscure associations between traits affecting dispersal ability and geographic range size of reef fishes in the Tropical Eastern Pacific

Alzate, Adriana; van der Plas, Fons; Zapata, Fernando A; Bonte, Dries; Etienne, Rampal S

Published in:
Ecology and Evolution

DOI:
[10.1002/ece3.4734](https://doi.org/10.1002/ece3.4734)

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2019

[Link to publication in University of Groningen/UMCG research database](#)

Citation for published version (APA):

Alzate, A., van der Plas, F., Zapata, F. A., Bonte, D., & Etienne, R. S. (2019). Incomplete datasets obscure associations between traits affecting dispersal ability and geographic range size of reef fishes in the Tropical Eastern Pacific. *Ecology and Evolution*, 9(4), 1567-1577. <https://doi.org/10.1002/ece3.4734>

Copyright

Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

Take-down policy

If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): <http://www.rug.nl/research/portal>. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

ORIGINAL RESEARCH

Incomplete datasets obscure associations between traits affecting dispersal ability and geographic range size of reef fishes in the Tropical Eastern Pacific

Adriana Alzate^{1,2,3}  | Fons van der Plas⁴  | Fernando A. Zapata⁵  | Dries Bonte²  | Rampal S. Etienne¹ 

¹Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands

²Terrestrial Ecology Unit, Ghent University, Ghent, Belgium

³German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Leipzig, Germany

⁴Institute for Special Botany and Functional Biodiversity, University of Leipzig, Leipzig, Germany

⁵Coral Reef Research Group, Department of Biology, Universidad del Valle, Cali, Colombia

Correspondence

Adriana Alzate, Groningen Institute for Evolutionary Life Sciences, University of Groningen, Groningen, The Netherlands.
Email: adria.alzate@gmail.com

Funding information

Nederlandse Organisatie voor Wetenschappelijk Onderzoek; Ubbo Emmius Fund; BelSpo IAP

Abstract

Dispersal is thought to be an important process determining range size, especially for species in highly spatially structured habitats, such as tropical reef fishes. Despite intensive research efforts, there is conflicting evidence about the role of dispersal in determining range size. We hypothesize that traits related to dispersal drive range sizes, but that complete and comprehensive datasets are essential for detecting relationships between species' dispersal ability and range size. We investigate the roles of six traits affecting several stages of dispersal (adult mobility, spawning mode, pelagic larval duration (PLD), body size, aggregation behavior, and circadian activity), in explaining range size variation of reef fishes in the Tropical Eastern Pacific (TEP). All traits, except for PLD (148 species), had data for all 497 species in the region. Using a series of statistical models, we investigated which traits were associated with large range sizes, when analyzing all TEP species or only species with PLD data. Furthermore, using null models, we analyzed whether the PLD-subset is representative of the regional species pool. Several traits affecting dispersal ability were strongly associated with range size, although these relationships could not be detected when using the PLD-subset. Pelagic spawners (allowing for passive egg dispersal) had on average 56% larger range sizes than nonpelagic spawners. Species with medium or high adult mobility had on average a 25% or 33% larger range, respectively, than species with low mobility. Null models showed that the PLD-subset was nonrepresentative of the regional species pool, explaining why model outcomes using the PLD-subset differed from the ones based on the complete dataset. Our results show that in the TEP, traits affecting dispersal ability are important in explaining range size variation. Using a regionally complete dataset was crucial for detecting the theoretically expected, but so far empirically unresolved, relationship between dispersal and range size.

KEYWORDS

adult mobility, aggregation behavior, body size, circadian activity, maximum linear distance, pelagic larval duration, spawning mode, tropical reef fishes

1 | INTRODUCTION

A key question in macroecology and biogeography is why there is so much variation in the geographic range sizes of species (Gaston, 2003). Several explanations have been suggested for this large variation: environmental and physical constraints, species differences in niche breadth, population abundance, latitudinal gradients, species' evolutionary age, body size, trophic level, colonization-extinction dynamics, and dispersal ability (reviewed in Gaston, 2003), all of which have the potential to interactively cause variation in range size. However, empirical studies investigating these explanations are scarce and their conclusions are often conflicting (Gaston, 2003). This is especially the case for tropical reef fishes, for which a general consensus on the principal determinants of their range sizes remains elusive in spite of much research effort (Ruttenberg & Lester, 2015).

Dispersal is one of the most obvious processes related to range expansion (Sexton, McIntyre, Angert, & Rice, 2009). It influences demography, colonization dynamics, local adaptation, speciation, and extinction (Holt & Gomulkiewicz, 1997; Hubbell, 2001; MacArthur & Wilson, 1967). Because reef fishes are usually confined to discrete, often isolated habitats, dispersal is expected to be a particularly strong determinant of range sizes (Leis, 1991; Victor, 1991). However, evidence for the existence of a relationship between dispersal ability and geographical range size in reef fishes is mixed at best (reviewed in Lester & Ruttenberg, 2005; Ruttenberg & Lester, 2015). Pelagic larval duration (PLD), which is the time period fish larvae spend in the water column prior to settlement, thus a proxy of dispersal (the longer the PLD the greater the dispersal ability), has shown to correlate poorly with range sizes, which has led others to question the importance of dispersal (Lester, Ruttenberg, Gaines, & Kinlan, 2007; Mora et al, 2011) and to suggest that other life-history traits are better predictors of range size (Luiz et al, 2013). Spawning mode (releasing either benthic or pelagic eggs) has been shown to be a good predictor of genetic structure in reef fishes (Riginos, Buckley, Blomberg, & Trembl, 2014). However, despite the fact that genetic structure should reflect the level of connectivity among populations (i.e., level of dispersal; Clobert, Danchin, Dhondt, & Nichols, 2001), spawning mode has not been found to be a good predictor of range size so far (Luiz et al, 2013).

Although it seems obvious that dispersal should play a key role in determining the geographical ranges of species, the question is why the effects of dispersal have rarely been demonstrated conclusively. This can be due to (at least) two reasons. Firstly, most of the studies have investigated the effects on range size of only one dispersive life stage: the larval stage (but see Luiz et al, 2013). This is usually justified by the assumption that reef fishes significantly disperse only during this period, when they can be transported by ocean currents. The time that larvae spend in the ocean before settlement, pelagic larval duration (PLD), has been the main studied trait affecting dispersal ability (Victor, 1991). Nevertheless, dispersal consists of several stages (departure, transfer, and settlement) and occurs during different life stages (Bonte et al, 2012). Overall dispersal ability is the result of the combined effect of multiple traits, which could have

evolved due to natural selection for high dispersal, or as an evolutionary by-product (Burgess, Baskett, Grosberg, Morgan, & Strathmann, 2015). Dispersal in reef fishes also occurs during the egg and adult life stages (Addis, Patterson, Dance, & Ingram, 2013; Appeldoorn, Hensley, Shapiro, Kioroglou, & Sanderson, 1994; Kaunda-Arara & Rose, 2004; Leis, 1978).

Although it is known that adults can move significant distances (Kaunda-Arara & Rose, 2004), the precise effect of dispersal during the adult life stage on range size has to the best of our knowledge not yet been studied. Adult body size can indirectly be related to dispersal via its relationship to fecundity. Larger females tend to have a higher fecundity than smaller females, thus allowing more eggs or larvae to potentially reach distant, suitable habitat (Thresher, 1984; Wootton, 1992). By increasing propagule pressure during range expansion, the chances of survival during the transfer stage might be higher, which might positively affect large-scale connectivity (Trembl et al, 2012). Aggregation behavior and nocturnal activity have been suggested to be related to the settlement stage of dispersal by decreasing predation risk, which increases colonization success (Luiz et al, 2013). Thus, various traits are directly or indirectly related to dispersal, and focusing on only one or very few of them might preclude the detection of links between dispersal ability and range sizes.

A second possible reason why it has been difficult to relate dispersal ability to range size is more methodological. Studies examining the range size–dispersal ability relationship in reef fishes have used only a subset of species (e.g., groups of species with known information on the trait of interest). The reason for this is that when investigating relationships between PLD and range size, scientists are limited by the relatively scarce availability of data. For instance, in the Tropical Eastern Pacific (TEP), this trait has been estimated for only 30% of the species. Studying a subset of species to test an ecological hypothesis is a common, more or less accepted practice in macroecology (Blackburn & Gaston, 1998). Nevertheless, an implicit assumption behind the use of species subsets is that this subset is representative of the total species pool, which may not always be the case. As pointed out by Blackburn and Gaston (1998), missing species do not only add noise to the macroecological patterns but might also distort them. Currently, we do not know the consequences of using only the subset of species for which PLD has been estimated when studying range sizes in tropical reef fishes.

Here, we tried to overcome some of the difficulties mentioned above by firstly investigating the effect of six dispersal-related traits (adult mobility, spawning mode, PLD, circadian activity, aggregation behavior, and body size) on range size, using only the subset of species for which PLD data is available (PLD-subset, 148 species). Secondly, we used null models to investigate whether the trait distribution for species in the PLD-subset is representative of the regional species pool. Then, we analyzed the complete dataset of reef fishes in the TEP (using five traits instead of six because of the need to exclude PLD data) to test whether (and if so, why) the results regarding the drivers of range sizes are different from those obtained when using only the PLD-subset. Finally, we

investigated how using an incomplete set of species may affect statistical model outcomes by analyzing 1,000 random subsets with the same number of species as the PLD-subset. We focus our study on the distribution of tropical reef fishes present in the TEP, which is a well-defined region with relatively clear limits, relatively isolated from other marine regions (such as from the Caribbean by the Panama Isthmus and from the Indo-Pacific by the large span of open ocean known as the East Pacific Barrier). The TEP marine fish fauna is well known, and information on geographic distribution and other species traits is available (Froese & Pauly, 2011; Robertson & Allen, 2016).

2 | METHODS

2.1 | Range size data

Data on geographic distribution of reef fish species were obtained from the Shorefishes of the Tropical Eastern Pacific (SFTEP), Online Information System (Robertson & Allen, 2016), and the IOBIS database (to complete the distributions of species that are not endemic to the TEP). We restricted our study to the distribution of tropical reef-associated bony fishes in the TEP region (*sensu* Robertson & Allen, 2016). We included only tropical species, whose main distribution occurs in the Pacific Ocean. For a total of 497 species, we calculated range size using the geographical coordinates of all records reported in the region. Range size was measured as the maximum linear distance (in kilometers) between any two points where a species has been recorded (Gaston, 1994). Range size was calculated using the function “geodist” from the R package “gmt” (Magnusson, 2014).

2.2 | Predictors of range size

We collated information on several species traits potentially affecting dispersal from the literature and online databases: body size, adult mobility, spawning mode, PLD, circadian activity, and aggregation behavior. All of these factors have been suggested as possible drivers of range size in reef fishes (e.g., Lester et al., 2007; Luiz et al., 2013; Ruttenberg & Lester, 2015).

Body size, the maximum recorded total length for each species, was obtained from Fish Base (Froese & Pauly, 2011) and Shore Fishes of the Tropical Eastern Pacific online information system—SFTEP (Robertson & Allen, 2016). Body size is related to many other life-history traits, also to habitat specialization and predation risk, which consequently could affect range size (Calder, 1984; Peters, 1983). Body size is positively related to fecundity (Thresher, 1984; Wootton, 1992; Zapata, 1990), increasing propagule pressure during range expansion and probably influencing large-scale connectivity (Tremblay et al., 2012). It is also positively related to adult mobility (Barlow, 1981) and home range size (Nash, Welsh, Graham, & Bellwood, 2015; Peters, 1983; Welsh & Bellwood, 2014), thereby potentially leading to larger range sizes (Gaston, 2003; Gaston & Blackburn, 1996).

Adult mobility was classified as low, medium, or high following Floeter, Ferreira, Dominici-Arosemena, and Zalmon (2004). Information for each species was collated from several studies (Appendix S1). Data on spawning mode were obtained from SFTEP (Robertson & Allen, 2016). Species were classified as pelagic or non-pelagic spawners (including species with benthic eggs, mouthbrooding and live birth). Pelagic spawners release their eggs in the water column, which are passively transported by water currents until the larvae hatch and are able to swim actively (Leis et al., 2013; Stobutzki, 1997). Thus, spawning mode gives an indication of dispersal in both the egg and larval stage (Leis, 2006; Leis et al., 2013). Data on pelagic larval duration (PLD) was obtained from the literature (Appendix S1). We supplemented these data with viviparous species, for which PLD = 0. Adult mobility, spawning mode, and PLD are all traits related to (but not actual measures of) dispersal ability that act at different life-history stages.

Data on circadian activity were obtained from Fish Base (Froese & Pauly, 2011), SFTEP (Robertson & Allen, 2016), as well as several other studies (Appendix S1). Species were classified as diurnal, crepuscular, or nocturnal. Data on aggregation behavior were obtained from Fish Base (Froese & Pauly, 2011), SFTEP (Robertson & Allen, 2016), and several other studies (Appendix S1). Species were classified as nonaggregative (species that never form any aggregation), temporarily aggregative (species that at some point in their lives form spawning or feeding aggregations), and aggregative (species that form aggregations or schools). Information on trophic level was collated from Fish Base (Froese & Pauly, 2011). Schooling (a form of aggregation) and nocturnal activity (a type of circadian activity) have been previously found to be good predictors of range size for tropical reef fishes (Luiz et al., 2013). They are suggested to reduce predation risk and increase the chances of survival and establishment after settlement (Luiz et al., 2013).

2.3 | Data analysis

2.3.1 | Analysis of the PLD-subset

We used general linear mixed models (LMMs) to study which factors explain variation in range size for reef-associated fish species in the TEP. We focused first on the species subset (148 species) for which PLD data is available. The model included six fixed factors, which are the species traits also described above: body size, spawning mode (pelagic or nonpelagic spawners), adult mobility (low, medium, or high), PLD, circadian activity (diurnal, crepuscular, or nocturnal), and aggregation behavior (temporarily aggregative, aggregative, nonaggregative). In addition, we controlled for possible phylogenetic effects (phylogenetic conservatism of range size) by including Genus, Family, and Order as nested random factors in the model. With an additional random factor (ocean basin), we controlled for the origin of the fauna: We distinguished between species that are endemic (TEP endemic) and species that are nonendemic to the TEP (e.g., transpacifics). Using ocean basin as a random factor follows similar approaches used in previous studies (Luiz et al., 2013). We included

this factor because transpacific species that reached the TEP have disproportionately larger range sizes than TEP endemics, which is due to differences in the largest distance between any two points within each ocean basin (~19,000 km vs. ~6,000 km).

We included a random factor that controls for the particular spatial structure of the TEP (a long continuous continental coastline + scattered oceanic islands), which imposes two different possible maximum range sizes (one for species that live on the coastline [~6,000 km] and one for the species living only on oceanic islands [~3,500 km]). The maximum linear distance between the northern and southern tips of the TEP is much longer than most of the island-mainland distances or island-island distances, potentially allowing species that range widely over continuous habitat to have broader geographic ranges than species that cross the gap between the islands and the mainland.

Predictors were tested for multicollinearity using the R code "HighstatLib.r" from Zuur, Ieno, and Elphick (2010). Variance inflation factors (VIF) of all variables were below 2.5, which is considered adequate for ensuring that variables are not collinear (Zuur et al., 2010). Range size data were logit-transformed to meet linear mixed model assumptions (Appendix S2).

The model was fitted using the "lmer" function from the R package "lme4" (Bates et al., 2014). We standardized the explanatory variables, in order to compare the effect sizes of the different predictors, using the function "standardize" from the R package "arm" (Gelman et al., 2009). We performed model selection using the function "dredge" from the R package "MuMIn" (Barton, 2013). To ensure robust model estimates, we used model averaging across all models (with an Akaike weight larger than 0.001) nested within the global model (Grueber,

Nakagawa, Laws, & Jamieson, 2011). Model averaging was done using the function "model.avg" from the R package "MuMIn" (Barton, 2013), which runs all possible models nested within the global model, calculates effect sizes of each predictor of each model, and calculates average effect sizes across models, weighed by Akaike weights.

We then used null models to test whether the PLD-subset is representative for the regional species pool, in terms of taxonomical, ecological, and life-history traits. We compared trait averages of 10,000 random subsets extracted from the complete dataset, each with 148 species (the same number of species as those with available PLD data), with the traits of the PLD-subset by calculating the difference in average trait values between the random and PLD-subset. Random subsets required that all levels per factor contained at least one data point (species). If average trait values of the PLD-subset were higher or lower than average trait values in 9,750 out of 10,000 random subsets, the difference was considered significant (two-sided test; $\alpha = 0.05$).

2.3.2 | Analysis of the complete dataset and random subsets

we repeated the general linear mixed models (LMMs) described above for the PLD data subset, but this time including all reef-associated fish species in the TEP. The only difference was that we could not include PLD as a predictor, due to data incompleteness.

As a sensitivity analysis, we repeated this model 1,000 times, but using random data subsets as described above. This way, we could investigate whether the analysis of random subsets of species, including species without PLD measurements, and without

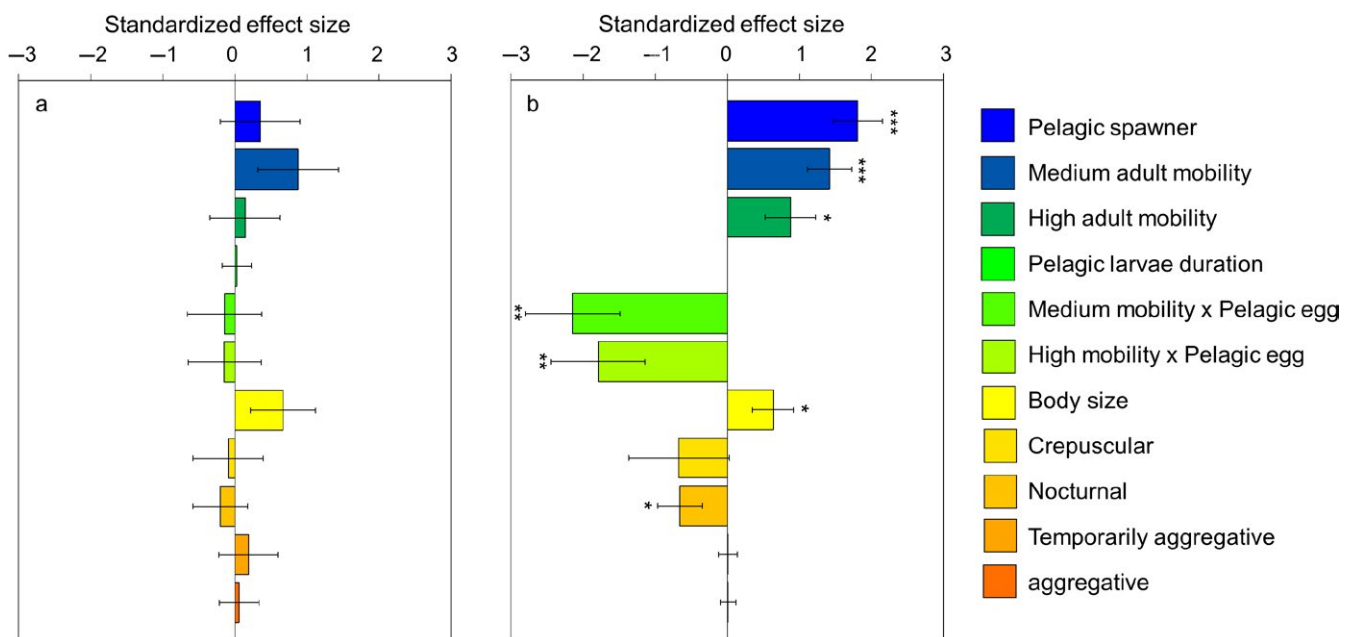


FIGURE 1 Standardized effect sizes of different predictors of range size for (a) the PLD-subset and (b) the complete dataset. Standard errors and significance levels (* < 0.05, ** < 0.001, *** < 0.0001) are shown. Reference levels: nonpelagic eggs, low adult mobility, nonaggregative, and diurnal

PLD as a predictor, yields the same results as the model for the PLD-subset. Random subsets required the presence of at least one data point (species) per level for each factor. The fitting and model selection procedures were the same as used for the PLD-subset.

3 | RESULTS

3.1 | Drivers of range size in the PLD-subset

When analyzing the drivers of range sizes using the PLD-subset, we found that none of the studied traits had a significant effect on range size (Figure 1a, Table 1 in Appendix S3). Null models indicated, however, that most trait values from the PLD-subset significantly differed from traits values obtained from randomly drawn species subsets (Figure 2). The PLD-subset contained, on average, a lower number of families (indicating a narrower phylogenetic breadth), a lower proportion of nonpelagic spawners, and a lower proportion of nonaggregative, nocturnal, and low mobility species than random subsets. The PLD-subset contained a higher proportion of pelagic spawners, and a higher proportion of aggregative, temporarily aggregative, diurnal, crepuscular, medium, and high mobility species than random subsets. The only trait that did not significantly differ between the PLD-subset and randomly drawn species subsets was adult body size.

3.2 | Predictors of range size in the complete TEP data and random data subsets

An analysis of the complete dataset, including all tropical species in the TEP, but excluding PLD from the model (because there is not

information for all species), revealed that the range size of reef fishes in the TEP is positively associated with several traits directly or indirectly related to dispersal ability (Figure 1b; Table 1 in Appendix S3). Spawning mode and adult mobility significantly interact to affect species range size (Figure 3 a,b, Table 1 in Appendix S3). While species that are pelagic spawners have larger ranges than nonpelagic spawners (by on average 5,070 km), the effect of adult mobility depends on the type of spawning mode (Figures 1 and 3 a,b). We found a strong positive relationship between adult mobility and range size for nonpelagic spawners, but not for pelagic spawners. Among nonpelagic spawners, species with high adult mobility have on average a range 6,829 km greater than that of species with low adult mobility. Body size and circadian behavior are also correlated with range size. Whilst body size has a positive effect on range size, nocturnal species attained significantly smaller ranges than diurnal species (Figure 1b, Table 1 in Appendix S3).

We performed the same statistical analysis for 1,000 random species subsets (with the same number of species as the PLD-subset). The effect sizes and significance levels (modal values) estimated from random species subsets were generally very similar to the ones estimated using the complete dataset (Figure 4). In contrast, effect sizes and significance levels of both the complete dataset and those of random subsets substantially differed from those of the PLD-subset for spawning mode and adult mobility: In the PLD-subset, positive effects of pelagic spawning mode and medium/high mobility on range size were underestimated (Figure 4). Furthermore, in the majority of cases, the *P* values by which spawning mode (88.9% of cases), medium (64.5% of cases), and high (89.7% of cases) adult mobility were related to range size were lower (i.e., more significant) in models based on random species subsets than in the PLD-subset,

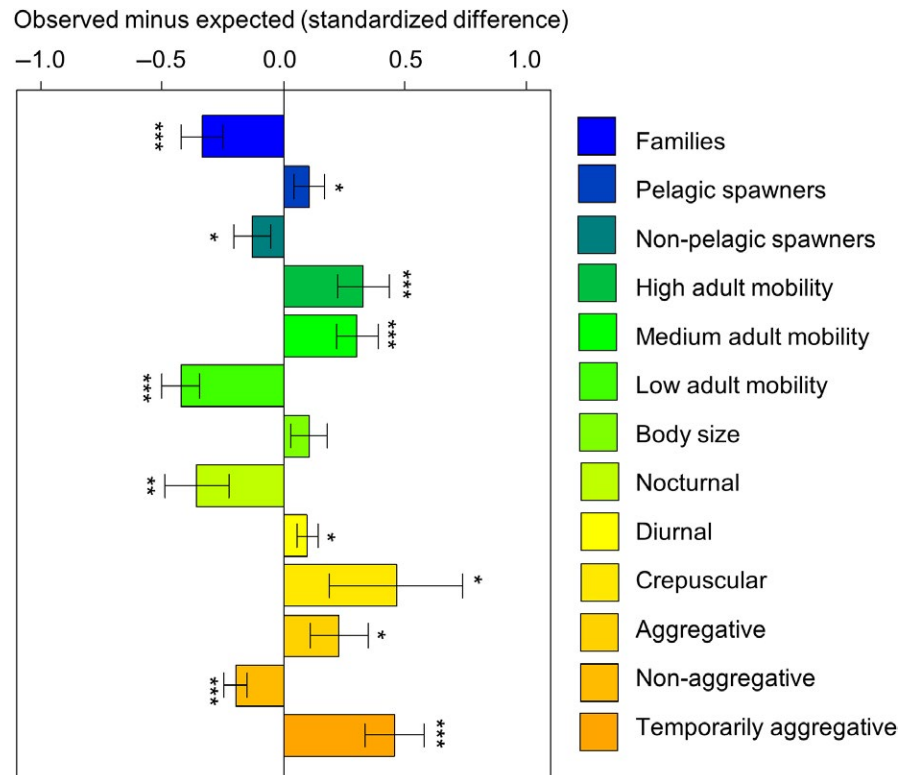


FIGURE 2 Standardized difference between the average trait value of 10,000 random species subsets and the PLD-subset. Both the random and the PLD-subset contained the same number of species (148). Positive values mean that the random subsets show higher values than the PLD-subset and vice versa for negative values. 95% confidence intervals and significance levels (* <0.05 , ** <0.001 , *** <0.0001) are shown

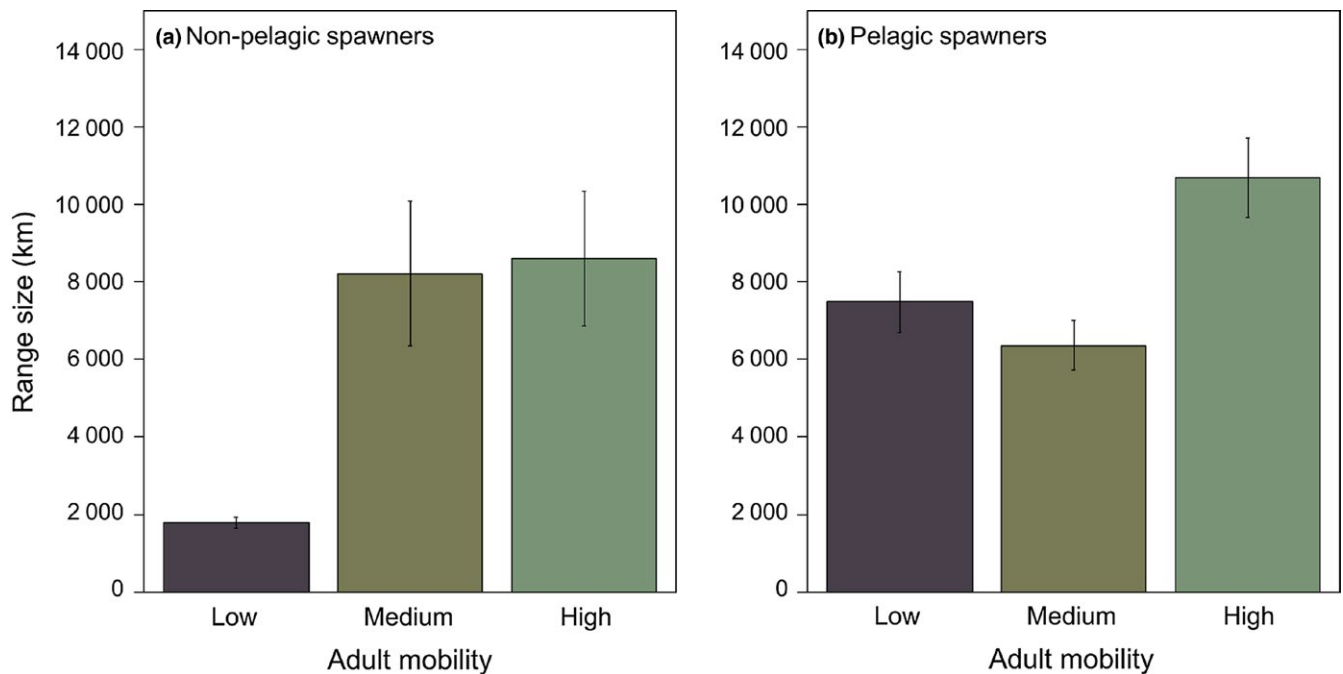


FIGURE 3 (a) range size significantly increases with adult mobility for species of nonpelagic spawners. (b) Range size is generally larger for pelagic spawners than for nonpelagic spawners, but does not significantly increase with adult mobility within the group of pelagic spawners. Standard errors are shown

even though sample sizes (and therefore statistical power) were equal (Figure 5).

4 | DISCUSSION

In spite of the obvious theoretical importance of dispersal as a potential determinant of range size (especially for organisms living in highly spatially clustered habitats), supporting empirical evidence has so far been surprisingly limited in both terrestrial (Dennis, Donato, Sparks, & Pollard, 2000; Edwards & Westoby, 1996; Gaston & Blackburn, 2003; Gutierrez & Menendez, 1997; Juliano, 1983; Kavanaugh, 1985; Malmqvist, 2000; McCauley, Davis, Werner, & Robeson, 2014) and marine organisms (Bonhomme & Planes, 2000; Brothers & Thresher, 1985; Hansen, 1980; Jablonski & Lutz, 1983; Jones, Caley, & Munday, 2002; Lester & Ruttenberg, 2005; Lester et al., 2007; Luiz et al., 2013; Mora et al., 2011; Mora, Chittaro, Sale, Kritzer, & Ludsin, 2003; Nanninga & Manica, 2018; Ruttenberg & Lester, 2015; Thresher & Brothers, 1985; Thresher, Colin, & Bell, 1989; Victor & Wellington, 2000; Wellington & Victor, 1989; Zapata & Herrón, 2002). Our study illustrates several reasons why it is difficult to find a definite answer to the question of whether or not dispersal is an important determinant of species range sizes. The first reason is the use of species subsets to tackle macroecological questions. We showed that when using the subset of species for which PLD data is available, none of the studied species traits significantly affected range size in the TEP. In contrast, when the dataset includes all species (but not necessarily all traits), several traits affecting dispersal ability are positively correlated with species

range sizes, at least in the TEP. A possible explanation is that the use of small data subsets does not allow to capture the true nature of the relationship between traits affecting dispersal ability and range size simply because of lack of statistical power. We tested this by analyzing random subsets of the TEP species pool and comparing the model outcomes with those based on the complete dataset and the PLD-subset. In contrast to analyses based on the PLD-subset, but in agreement with the analysis based on the complete dataset, most models based on random species subsets indicated that both spawning mode and adult mobility are significantly positively correlated with range size. Thus, the lack of relationship between range size and traits related to dispersal ability is not likely caused by the use of smaller datasets per se, but by the particular nature of the PLD-subset, which consists of a highly nonrandom set of species.

The fact that PLD has only been estimated for 30% of the reef-associated species in the TEP (497 in total) and that 32% of these species come from only three families (Labridae =13.6%, Pomacentridae =10.7%, and Serranidae =7.9%) could have been a reason for why we did not detect any signal in the analysis of the PLD species subset. We showed, using null models, that the PLD-subset is not a representative sample from the regional species pool: 12 out of 13 characteristics that we compared differed significantly between the PLD-subset and random species subsets. Therefore, the PLD-subset for the TEP is biased in terms of taxonomical breadth and species traits. It contains a smaller number of families, a lower proportion of nocturnal, nonpelagic and low mobile species and a higher proportion of diurnal, aggregative and more dispersive species (pelagic spawners with high/medium adult mobility) than expected based on random species subsampling. Potential problems related to the use of biased datasets

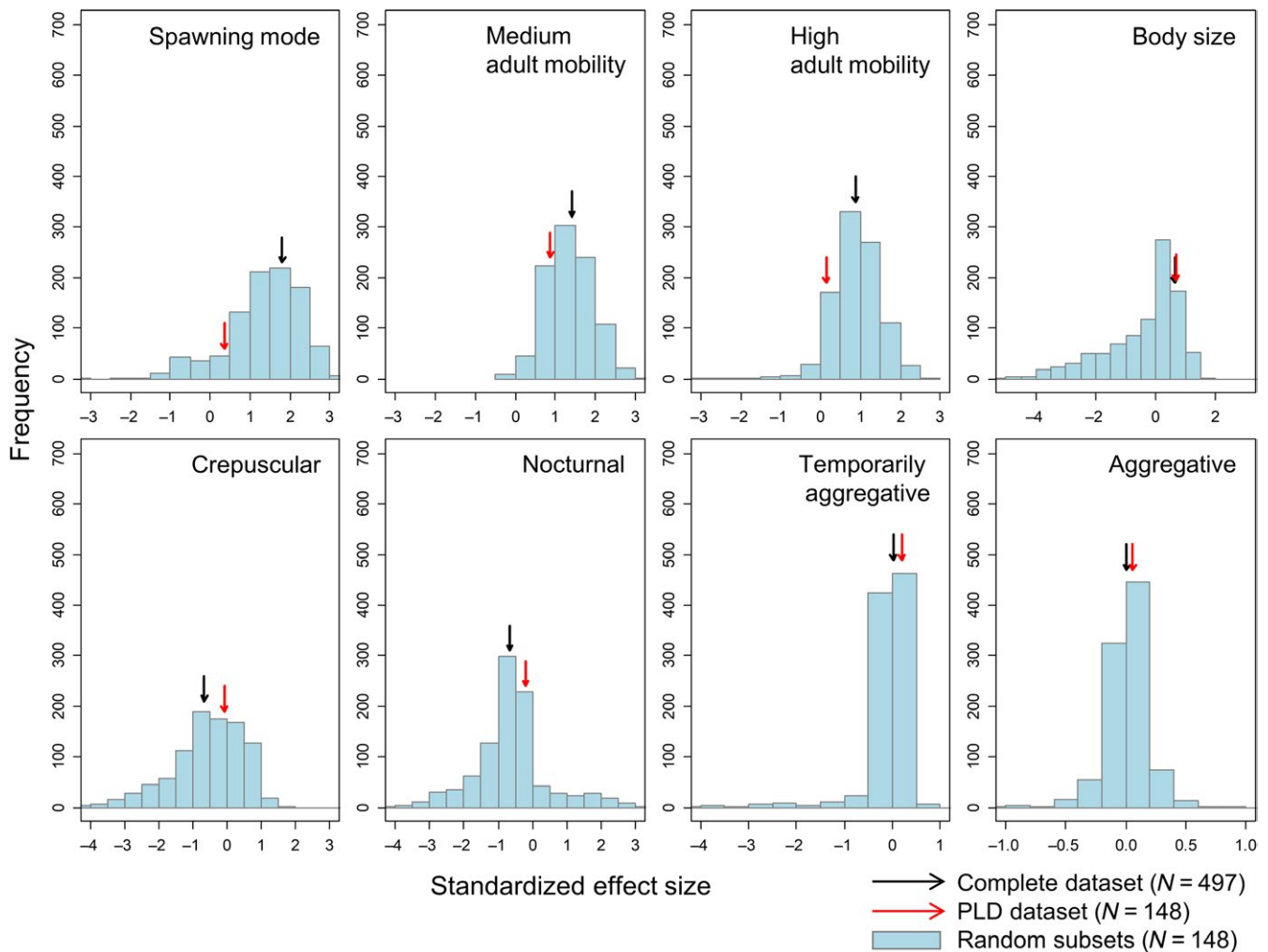


FIGURE 4 Frequency distributions of effect sizes of model coefficients for 1,000 linear mixed model analyses of equal number of random species subsets. The black and red arrows show the effect size values for the complete dataset and the PLD-subset, respectively

have previously been discussed verbally (Blackburn & Gaston, 1998). Our study quantitatively shows that these problems are far from trivial and that biased datasets can have severe implications for the outcomes and conclusions drawn from a study. A similar example of such an effect is shown in a study examining how species traits influence extinction risk in mammals, where a seemingly strong relationship between body mass and extinction risk (larger mammals being more vulnerable) is actually driven by the biased availability of data, where large (usually rare) mammals have been generally better studied than small ones (Gonzalez-Suarez, Lucas, & Revilla, 2012). Therefore, general conclusions based on the analysis using the PLD-subset should be taken with extreme caution, because the nonrandom sample from the regional species pool may bias model estimates (Little & Rubin, 2002; Nakagawa & Freckleton, 2008).

A second reason why we found, in contrast to many studies, a clear relationship between dispersal and range size, is that most studies have focused on single or few dispersal-related traits, thereby (implicitly) denying the fact that dispersal is more complex including traits related to departure, transfer, and settlement decisions. The inclusion of other traits likely to affect dispersal ability (e.g.,

spawning mode, body size, adult mobility) revealed that several of these traits, some of which are typically not studied, are strong predictors of range sizes. Our analyses of the complete dataset, as well as those based on random subsets, showed that several traits related to different stages of dispersal (spawning mode, body size, and adult mobility) correlate positively with range size in the TEP. Species that are pelagic spawners that are more mobile during their adult stage and that attain larger body sizes have larger ranges than nonpelagic spawner, less mobile, and smaller-bodied species. The effect of adult mobility on range size depends on the type of spawning mode. While adult mobility does not affect range size of pelagic spawners, it does significantly affect the range size of nonpelagic spawners. In this study, we could not detect significant relationships between PLD and range sizes. Similarly, another study has shown that larval swimming capacity is a better proxy of dispersal than PLD per se (Nanninga & Manica, 2018). Nevertheless, the importance of PLD for range size cannot be completely discarded until more complete information on this trait is available for a larger number of species.

Pelagic spawners are able to attain larger ranges than nonpelagic spawners because pelagic eggs spend some time in the water column,

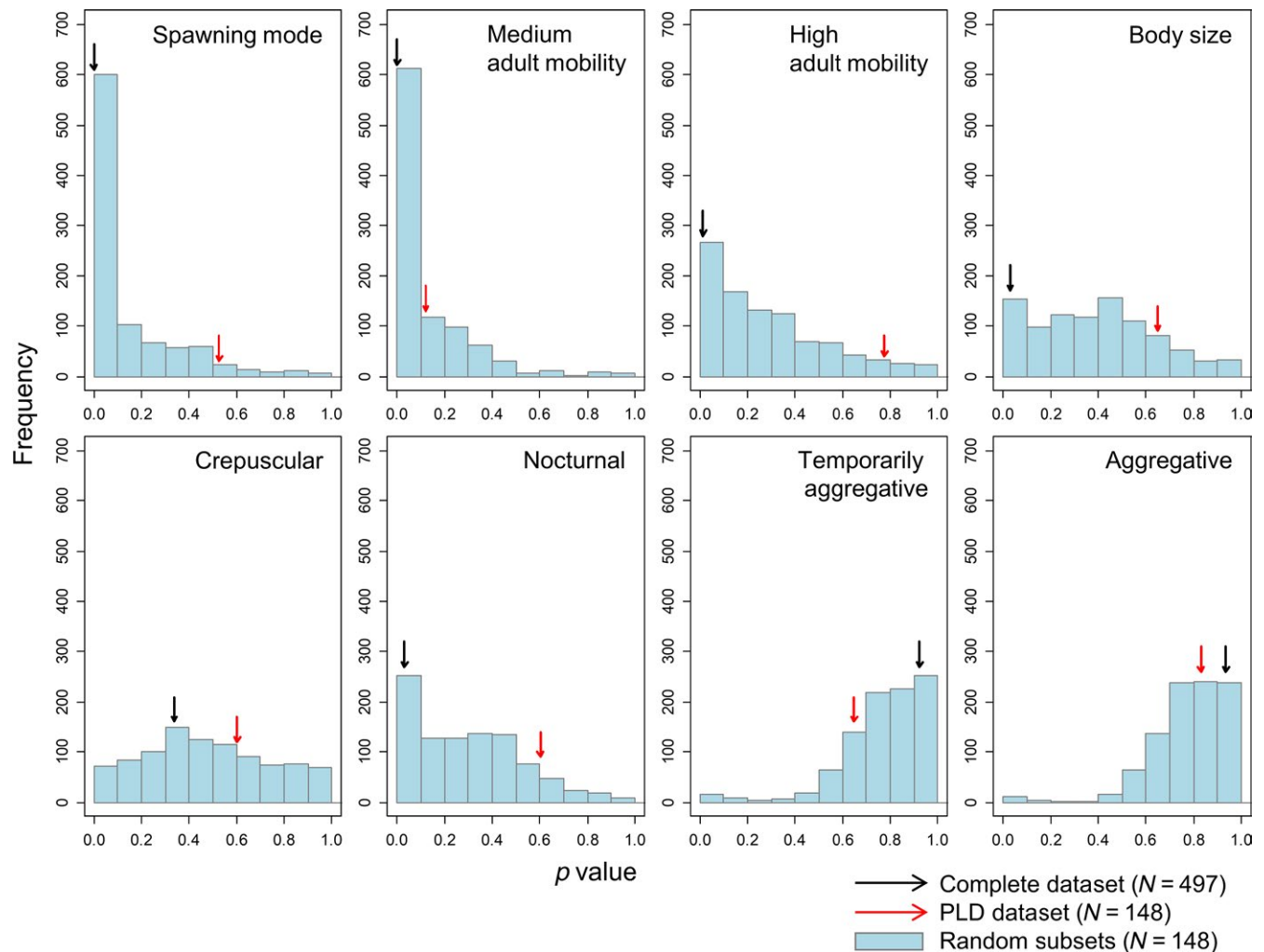


FIGURE 5 Frequency distributions for the significance of model coefficients for 1,000 random species subsets. The black and red arrows show the significance values for the complete dataset and the PLD-subset, respectively

which allows for passive dispersal without experiencing the energetic costs that larvae do (Leis et al, 2013; Stobutzki, 1997). Larvae of non-pelagic spawners hatch at a more developed stage and are more developed at hatching. Thus, have better control of their dispersal than those of pelagic spawners (Leis et al, 2013; Wootton, 1992). More mature larvae become more active swimmers and are less likely to passively disperse for long distances through ocean currents (Leis, 1991, 2006; Leis et al, 2013; Munday & Jones, 1998; Stobutzki & Bellwood, 1997). Therefore, passive dispersal is expected to be the predominant dispersal mode at least at the beginning of the larval stage for pelagic spawners, whereas active dispersal is likely to be predominant for nonpelagic spawners earlier in their larval life.

The positive effect of body size on range size is likely to be more indirect, as body size can affect fecundity, dispersal during the egg and larval stage, survival, and adult mobility. Larger females usually have higher fecundities, which means more propagules and therefore a higher probability that at least some of those propagules (larvae) will survive and colonize distant localities (Thresher, 1984; Wootton, 1992; Beldade et al, 2012). An increase in propagule pressure during range expansion can have consequences for large-scale connectivity (Trembl

et al, 2012). In addition, larger species tend to produce smaller eggs which are more likely to be dispersed passively (Thresher, 1984). Body size is positively related to adult mobility (Barlow, 1981) and home range size (Welsh & Bellwood, 2014), thereby potentially leading to larger range sizes (Gaston, 2003; Gaston & Blackburn, 1996). In contrast to another study based on all ocean basins (Luiz et al, 2013), our study did not find relationships between range size and aggregation behavior. This trait has been treated as proxies for predation avoidance and subsequent successful colonization of new habitat patches (Luiz et al, 2013), but our study cannot provide evidence for the importance of this factor in explaining range sizes in the TEP. We also found a negative relationship between nocturnality and range sizes. This was an unexpected result and in contrast to previous findings (Luiz et al, 2013).

Another possible reason why our results contrast with other studies that did not find strong links between traits that affect dispersal ability and range size is that mechanisms underlying range expansion in the TEP might differ from mechanisms that predominate in other regions. First, all oceans basins differ in size, shape, geographical barriers, habitat availability, and configuration, which can influence the maximum range that species can attain (Ruttenberg & Lester, 2015).

In the TEP, the reef habitat available for reef fishes is mostly distributed along the coast in a semicontinuous manner (except for two wide stretches of sand that act as dispersal barriers or filters) and around the few oceanic islands. Such extrinsic factors could not only influence range size, but also which and how species traits influence range expansion. Certainly, other factors, not studied in detail here, such as species age and geological history (e.g., habitat suitability and isolation; Pellissier et al, 2014; Ruttenberg & Lester, 2015) also might drive variation in range sizes, and it is likely that these factors interact with dispersal. Hence, future studies investigating these factors simultaneously, using (regionally) complete datasets, might provide even greater understanding in the distribution of range sizes.

In summary, we showed that several traits presumed to affect dispersal ability are important determinants of range size and that using several of these traits and a regionally complete dataset, rather than a subset of the species in a region, was crucial for reaching these conclusions. Our study serves as a warning about the crucial importance of choosing an adequate dataset to study macroecological patterns. We demonstrated that the use of a species subset, which is not a random sample from the regional species pool (the PLD-subset), can produce very different results than the complete dataset or subsets based on random samples.

ACKNOWLEDGMENTS

We thank two anonymous reviewers for useful comments that helped improve our manuscript. AA was funded by the Ubbo Emmius Fund and by BelSpo IAP project "SPatial and environmental determinants of Eco-Evolutionary DYnamics: anthropogenic environments as a model." RSE thanks the Netherlands Organization for Scientific Research (NWO) for financial support through a VICI grant.

AUTHORS' CONTRIBUTION

AA developed the research idea and collated the data. AA and FvdP performed the statistical analyses. AA wrote the first draft of the manuscript. AA, FvdP, FAZ, DB, and RSE contributed with discussions and comments on the manuscript.

DATA ACCESSIBILITY

Data on range size and traits of reef fishes used in this study are available on the Dryad digital repository <https://doi.org/10.5061/dryad.ns5m4kg>.

ORCID

Adriana Alzate  <https://orcid.org/0000-0002-8236-6510>

Fons van der Plas  <https://orcid.org/0000-0003-4680-543X>

Fernando A. Zapata  <https://orcid.org/0000-0001-8989-4458>

Dries Bonte  <https://orcid.org/0000-0002-3320-7505>

Rampal S. Etienne  <https://orcid.org/0000-0003-2142-7612>

REFERENCES

- Addis, D. T., Patterson, W. F. III, Dance, M. A., & Ingram, G. W. Jr (2013). Implication of reef fish movement from unreported artificial reef sites in the northern Gulf of Mexico. *Fisheries Research*, 147, 349–358.
- Appeldoorn, R. S., Hensley, D. A., Shapiro, D. G., Kioroglou, S., & Sanderson, B. G. (1994). Egg dispersal in a Caribbean coral reef fish, *Thalassoma bifasciatum*. II. Dispersal off the reef platform. *Bulletin of Marine Science*, 54(1), 271–280.
- Barlow, G. W. (1981). Patterns of parental investment, dispersal and size among coral-reef fishes. *Environmental Biology of Fishes*, 6(1), 65–85.
- Barton, K. (2013). MuMIn: Multi-model inference. R package version 1.9.5. <http://CRAN.R-project.org/package=MuMIn>. Accessed 2014.
- Bates, D., Maechler, M., Bolker, B., Walker, S., Bojesen Christensen, R. H., & Singmann, H.. (2014) Linear mixed-effects models using Eigen and S4. R package version 1.1-6. <http://lme4.r-forge.r-project.org/package=lme4>.
- Blackburn, T. M., & Gaston, K. J. (1998). Some methodological issues in macroecology. *The American Naturalist*, 151(1), 68–83. <https://doi.org/10.1086/286103>
- Beldade, R., Holbrook, S. J., Schmitt, R. J., Planes, S., Malone, D., & Bernardi, G. (2012). Larger female fish contribute disproportionately more to self-replenishment. *Proceedings of the Royal Society B*, 279, 2116–2121.
- Bonhomme, F., & Planes, S. (2000). Some evolutionary arguments about what maintains the pelagic larvae in reef fishes. *Environmental Biology of Fishes*, 59, 365–383.
- Bonte, D., Van Dyck, H., Bullock, J. M., Coulon, A., Delgado, M., Gibbs, M., ... Travis, J. M. J. (2012). Costs of dispersal. *Biological reviews of the Cambridge Philosophical Society*, 87, 290–312.
- Brothers, E. B., & Thresher, R. E.. (1985) Pelagic duration, dispersal, and the distribution of Indo-Pacific coral-reef fishes. The ecology of coral reefs, ed Reaka ML (Symposia Series for Undersea Research, NOAA's Undersea Research Program, Vol. 3, No. 1), pp 53–69.
- Burgess, S. C., Baskett, M. L., Grosberg, R. K., Morgan, S. G., & Strathmann, R. R. (2015). When is dispersal for dispersal? Unifying marine and terrestrial perspectives. *Biological Reviews*, 91(3), 867–882.
- Calder, W. A. (1984). *Size, Function and life history* (pp. 450). Mineola, New York: Dover Publications, INC.
- Clobert, J., Danchin, E., Dhondt, A. A., & Nichols, J. D. (2001). *Dispersal* (pp. 474). New York, NY: Oxford University Press.
- Dennis, R. L. H., Donato, B., Sparks, T. H., & Pollard, E. (2000). Ecological correlates of island incidence and geographical range among British butterflies. *Biodiversity and Conservation*, 9, 343–359.
- Edwards, W., & Westoby, M. (1996). Reserve mass and dispersal investment in relation to geographic range of plant species: Phylogenetically independent contrast. *Journal of Biogeography*, 23, 329–338.
- Floeter, S. R., Ferreira, C. E. L., Dominici-Arosemena, A., & Zalmon, I. R. (2004). Latitudinal gradients in Atlantic reef fish communities: Trophic structure and spatial use patterns. *Journal of Fish Biology*, 64, 1680–1699. <https://doi.org/10.1111/j.0022-1112.2004.00428.x>
- Froese, R., & Pauly, D. P.(2011). FishBase. Accessed 2013. World Wide Web electronic publication. www.fishbase.org.
- Gaston, K. J. (1994). Measuring geographic range sizes. *Ecography*, 17(2), 198–205. <https://doi.org/10.1111/j.1600-0587.1994.tb00094.x>
- Gaston, K. J. (2003). *The structure and dynamics of geographic ranges*. Oxford, UK: Oxford University Press.
- Gaston, K. J., & Blackburn, T. M. (1996). Range size-body size relationships: Evidence of scale dependence. *Oikos*, 75, 479–485. <https://doi.org/10.2307/3545889>
- Gaston, K. J., & Blackburn, T. M. (2003). Dispersal and the interspecific abundance-occupancy relationship in British bird. *Global Ecology and Biogeography*, 12(5), 373–379.
- Gelman, A., Su, Y.-S., Yajima, M., Hill, J., Pittau, M. G., Kerman, J., & Zheng, T. (2009). arm: Data analysis using regression and

- multilevel/hierarchical models. R Package, Version, 9, 01. <https://CRAN.R-project.org/package=arm>.
- Gonzalez-Suarez, M., Lucas, P. M., & Revilla, E. (2012). Biases in comparative analyses of extinction risk: Mind the gap. *Journal of Animal Ecology*, 81, 1211–1222. <https://doi.org/10.1111/j.1365-2656.2012.01999.x>
- Grueber, C. E., Nakagawa, S., Laws, R. J., & Jamieson, I. G. (2011). Multimodel inference in ecology and evolution: Challenges and solutions. *Journal of Evolutionary Biology*, 24, 699–711. <https://doi.org/10.1111/j.1420-9101.2010.02210.x>
- Gutierrez, D., & Menendez, R. (1997). Patterns in the distribution, abundance and body size of carabid beetles (Coleoptera: Caraboidea) in relation to dispersal ability. *Journal of Biogeography*, 24, 903–914. <https://doi.org/10.1046/j.1365-2699.1997.00144.x>
- Hansen, T. H. (1980). Influence of larval dispersal and geographic distribution on species longevities in neogastropods. *Paleobiology*, 6, 193–207.
- Holt, R. D., & Gomulkiewicz, R. (1997). How does immigration influence local adaptation? A re-examination of a familiar paradigm. *The American Naturalist*, 149, 563–572. <https://doi.org/10.1086/286005>
- Hubbell, S. P. (2001). *The Unified Neutral Theory of Biodiversity and Biogeography*. Princeton, NJ: Princeton University Press.
- Jablonski, D., & Lutz, R. A. (1983). Larval ecology of marine benthic invertebrates: Paleobiological implications. *Biological Reviews*, 58, 21–89. <https://doi.org/10.1111/j.1469-185X.1983.tb00380.x>
- Jones, G. P., Caley, M. J., & Munday, P. L. (2002). Rarity in coral reef fish communities. In P. F. Sale (Ed.), *Coral reef fishes: Dynamics and diversity in a complex ecosystem* (pp. 81–101). San Diego, CA: Academic Press.
- Juliano, S. A. (1983). Body size, dispersal ability, and range size in North American species of *Brachinus* (Coleoptera: Carabidae). *Coleopterist Bulletin*, 37, 232–238.
- Kaunda-Arara, B., & Rose, G. A. (2004). Long-distance movements of coral reef fishes. *Coral Reefs*, 23, 410–412. <https://doi.org/10.1007/s00338-004-0409-7>
- Kavanaugh, D. H. (1985). On wing atrophy in carabid beetles (Coleoptera: Carabidae), with special reference to Nearctic *Nebria*. In G. E. Ball (Ed.), *Taxonomy, Phylogeny and zoogeography of beetles and ants* (pp. 408–431). Dordrecht, Netherlands: Junk.
- Leis, J. M. (1978). Systematics and zoogeography of the porcupine fishes (Diodon, Diodontidae, Tetraodontiformes), with comments on egg and larval development. *Fishery Bulletin*, 76(3), 535–567.
- Leis, J. M. (1991). The pelagic stage of reef fishes: The larval biology of coral reef fishes. In P. F. Sale (Ed.), *The Ecology of Fishes on Coral Reefs* (pp. 183–230). San Diego, CA: Academic Press.
- Leis, J. M. (2006). Are larvae of demersal fishes plankton or nekton? *Advances in Marine Biology*, 51, 58–141.
- Leis, J. M., Caselle, J. E., Bradbury, I. R., Kristiansen, T., Llopis, J. K., Miller, M. J., ... Warner, R. R. (2013). Does fish larval dispersal differ between high and low latitudes? *Proceedings of the Royal Society B*, 280, 20130327. <https://doi.org/10.1098/rspb.2013.0327>
- Lester, S. E., & Ruttenberg, B. I. (2005). The relationship between pelagic larval duration and range size in tropical reef fishes: A synthetic analysis. *Proceedings of the Royal Society B*, 72, 585–591. <https://doi.org/10.1098/rspb.2004.2985>
- Lester, S. E., Ruttenberg, B. I., Gaines, S. D., & Kinlan, B. P. (2007). The relationship between dispersal ability and geographic range size. *Ecology Letters*, 10, 745–758. <https://doi.org/10.1111/j.1461-0248.2007.01070.x>
- Little, R. J. A., & Rubin, D. B. (2002). *Statistical analysis with missing data* (pp. 408). New York, NY: John Wiley & Sons.
- Luiz, O. J., Allen, A. P., Robertson, D. R., Floeter, S. R., Kulbicki, M., Vigliola, L., ... Madin, J. S. (2013). Adult and larval traits as determinants of geographic range size among tropical reef fishes. *Proceedings of the National Academy of Sciences*, 110(41), 16498–16502. <https://doi.org/10.1073/pnas.1304074110>
- MacArthur, R. H., & Wilson, E. O. (1967). *The theory of island biogeography*. Princeton, NJ: Princeton University Press.
- Magnusson, A. (2014) *gmt: Interface between GMT map-making software and R*. R package version 1.2-0. <http://CRAN.R-project.org/package=gmt>.
- Malmqvist, B. (2000). How does wing length relate to distribution patterns of stoneflies (Plecoptera) and mayflies (Ephemeroptera)? *Biological Conservation*, 93, 271–276. [https://doi.org/10.1016/S0006-3207\(99\)00139-1](https://doi.org/10.1016/S0006-3207(99)00139-1)
- McCauley, S. J., Davis, C. J., Werner, E. E., & Robeson, M. S. II (2014). Dispersal, niche breadth and population extinction: Colonization rations predict range size in North American dragonflies. *Journal of Animal Ecology*, 83, 858–865.
- Mora, C., Chittaro, P. M., Sale, P. F., Kritzer, J. P., & Ludsins, S. A. (2003). Patterns and processes in reef fish diversity. *Nature*, 421, 933–936. <https://doi.org/10.1038/nature01393>
- Mora, C., Treml, E., Robert, J., Crosby, K., Roy, D., & Titterson, D. P. (2011). High connectivity among habitats precludes the relationship between dispersal and range size in tropical reef fishes. *Ecography*, 35(1), 89–96.
- Munday, P. L., & Jones, G. P. (1998). The ecological implications of small body size among coral-reef fishes. *Oceanography and Marine Biology: an Annual Review*, 36, 373–411.
- Nakagawa, S., & Freckleton, R. P. (2008). Missing inaction: The dangers of ignoring missing data. *Trends in Ecology and Evolution*, 23, 592–596. <https://doi.org/10.1016/j.tree.2008.06.014>
- Nanninga, G. B., & Manica, A. (2018). Larval swimming capacities affect genetic differentiation and range size in demersal marine fishes. *Marine Ecology Progress Series*, 589, 1–12. <https://doi.org/10.3354/meps12515>
- Nash, K. L., Welsh, J. Q., Graham, N. A., & Bellwood, D. R. (2015). Home-range allometry in coral reef fishes: Comparison to other vertebrates, methodological issues and management implications. *Oecologia*, 177(1), 73–83. <https://doi.org/10.1007/s00442-014-3152-y>
- Pellissier, L., Leprieux, F., Parravicini, V., Cowman, P. F., Kulbicki, M., Litsios, G., ... Mouillot, D. (2014). Quaternary coral reef refugia preserved fish diversity. *Science*, 344, 1016–1019. <https://doi.org/10.1126/science.1249853>
- Peters, R. H. (1983). *The ecological implications of body size*. Cambridge, UK: Cambridge University Press.
- Riginos, C., Buckley, Y. M., Blomberg, S. P., & Treml, E. A. (2014). Dispersal capacity predicts both population genetic structure and species richness in reef fishes. *The American Naturalist*, 148, 52–64. <https://doi.org/10.1086/676505>
- Robertson, D. R., & Allen, G. (2016). *Fishes: East Pacific. An identification guide to the shore-fish fauna of the tropical Eastern Pacific*. Santa Cruz, CA: Copyright Smithsonian Institution, Left Coast R&C. <http://biogeodb.stri.si.edu/bioinformatics/sfstep/taxon.php>.
- Ruttenberg, B. I., & Lester, S. E. (2015). Patterns and processes in geographic range size in coral reef fishes. In C. Mora (Ed.), *Ecology of fishes on coral reefs* (pp. 97–103). Cambridge, UK: Cambridge University Press.
- Sexton, J. P., McIntyre, P. J., Angert, A. L., & Rice, K. J. (2009). Evolution and ecology of species range limits. *Annual Review of Ecology, Evolution and Systematics*, 40, 415–436. <https://doi.org/10.1146/annurev.ecolsys.110308.120317>
- Stobutzki, I. C. (1997). Energetic cost of sustained swimming in the late pelagic stages of reef fishes. *Marine Ecology Progress Series*, 152, 249–259. <https://doi.org/10.3354/meps152249>
- Stobutzki, I. C., & Bellwood, D. R. (1997). Sustained swimming abilities of the late pelagic stages of coral reef fishes. *Marine Ecology Progress Series*, 149, 35–41. <https://doi.org/10.3354/meps149035>
- Thresher, R. E. (1984). *Reproduction in reef fishes*. Neptune City, NJ: TFH Publications.

- Thresher, R. E., & Brothers, E. B. (1985). Reproductive ecology and biogeography of Indo-West Pacific angelfishes (Pisces: Pomacanthidae). *Evolution*, 39, 878–887. <https://doi.org/10.1111/j.1558-5646.1985.tb00429.x>
- Thresher, R. E., Colin, P. L., & Bell, L. J. (1989). Planktonic duration, distribution and population structure of western and central Pacific damselfishes (Pomacentridae). *Copeia*, 2, 420–434. <https://doi.org/10.2307/1445439>
- Trembl, E. A., Roberts, J. J., Chao, Y., Halpin, P. N., Possingham, H. P., & Riginos, C. (2012). Reproductive output and duration of the pelagic larval stage determine seascape-wide connectivity of marine populations. *Integrative and Comparative Biology*, 52, 525–537. <https://doi.org/10.1093/icb/ics101>
- Victor, B. C. (1991). Settlement strategies and biogeography of reef fishes. In P. F. Sale (Ed.), *The ecology of fishes on coral reefs* (pp. 231–260). San Diego, CA: Academic Press.
- Victor, B. C., & Wellington, G. M. (2000). Endemism and the pelagic larval duration of reef fishes in the eastern Pacific Ocean. *Marine Ecology Progress Series*, 205, 241–248. <https://doi.org/10.3354/meps205241>
- Wellington, G. M., & Victor, B. C. (1989). Planktonic larval duration of one hundred species of Pacific and Atlantic damselfishes (Pomacentridae). *Marine Biology*, 101, 557–568. <https://doi.org/10.1007/BF00541659>
- Welsh, J. Q., & Bellwood, D. R. (2014). Herbivorous fishes, ecosystem function and mobile links on coral reefs. *Coral Reefs*, 33, 303–311. <https://doi.org/10.1007/s00338-014-1124-7>
- Wootton, R. J. (1992). *Fish Ecology*. New York, NY: Chapman & Hall.
- Zapata, F. A. (1990). Local and regional patterns of distribution and abundance in marine reef fishes. Tucson, AZ: The University of Arizona. 102–150. PhD thesis.
- Zapata, F. A., & Herrón, P. A. (2002). Pelagic larval duration and geographic distribution of tropical eastern Pacific snappers (Pisces: Lutjanidae). *Marine Ecology Progress Series*, 230, 295–300. <https://doi.org/10.3354/meps230295>
- Zuur, A. F., Ieno, E. N., & Elphick, C. S. (2010). A protocol for data exploration to avoid common statistical problems. *Methods in Ecology and Evolution*, 1, 3–14. <https://doi.org/10.1111/j.2041-210X.2009.00001.x>

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Alzate A, van der Plas F, Zapata FA, Bonte D, Etienne RS. Incomplete datasets obscure associations between traits affecting dispersal ability and geographic range size of reef fishes in the Tropical Eastern Pacific. *Ecol Evol*. 2019;9:1567–1577. <https://doi.org/10.1002/ece3.4734>